

**Statistics Final Exam Review Topics****I. Probability rules**

Ex:

- In 2009, a Gallup poll found that 47% of Americans favored protection of the environment at the risk of limiting energy supplies, while 46% favored the development of U.S. energy supplies even if that risked harming the environment. Another 3% felt that energy and the environment were equally important. Based on this poll, what is the probability a person had “no opinion”?
- Suppose we select two people at random from this poll. What is the probability that both favored the environment?
- What is the probability that neither favored the environment?
- What does mutually exclusive mean?
- A consumer organization estimates that over a 1-year period, 17% of cars will need to be repaired once, 7% of cars will need repairs twice, and 4% will require three or more repairs. What is the probability that a car chosen at random will need
  - no repairs?
  - no more than one repair?
  - some repairs?
- You roll a fair die three times. What is the probability that
  - you roll all 6's?
  - you roll all odd numbers?
  - None of your rolls gets a number divisible by 3?
  - You roll at least one 5?
  - The numbers you roll are not all 5s?
- Use the table below to answer the questions that follow:

		Goals			
Sex		Grades	Friendship	Sports	Total
	Boy	117	50	60	227
	Girl	130	91	30	251
	Total	247	141	90	478

- What is the probability that a person's goal is sports?
- What is the probability that a person's goal is sports given that she is a girl?
- Are “Sports” and “Girl” independent?

**2. Making & using Venn diagrams**

Ex:

- Employment data at a large company reveal that 72% of workers are married, 44% are college graduates, and that half of the college graduates are married. What is the probability that a randomly-selected worker:
  - Is neither married nor a college graduate?
  - Is married but not a college graduate?
  - Is married **or** a college graduate?
- 56% of all American workers have a workplace retirement plan. 68% have health insurance, and 49% have both benefits. We select a worker at random.
  - What is the probability that he has neither employer-sponsored health insurance nor a retirement plan?
  - What's the probability he has health insurance if he has a retirement plan?
  - Are having health insurance and a retirement plan independent events?
  - Are having health insurance and a retirement plan mutually exclusive events?

### 3. Making & using tree diagrams

Ex:

- A recent Maryland highway safety study found that in 77% of all accidents the driver was wearing a seat belt. Accident reports indicated that 92% of those drivers escaped serious injury but only 63% of non-belted drivers were so fortunate. Overall, what's the probability that a driver involved in an accident was seriously injured? Make a tree diagram to answer this question.
- A private college report contains these statistics:  
70% of incoming freshmen attended public schools. 75% of public school students who enroll as freshmen eventually graduate. 90% of other freshmen eventually graduate. Based on these probabilities, what percent of freshmen eventually graduate? Make a tree diagram to answer this question.

### 4. Finding the median, IQR, minimum, maximum, Q1, Q3

Ex:

Ten neighborhood kids went out to Trick or Treat on Halloween night. Here is a list of the number of treats they collected:

45, 34, 56, 32, 10, 32, 62, 11, 55, 34

- Find the mean and median number of treats.
- Find the Q1, Q3, and IQR of the number of treats.
- Suppose the kid who first came back with 62 treats went back out. At the end of the night, he ended up with 262 treats. Find the new mean and median for these ten children.
- How do the new mean and median compare to the original values?
- Which does a better job of describing the typical number of treats for the new data – the mean or median? Why?
- The kid who got only 10 treats got 6 small candy bars weighing 2oz. each, 3 large bars weighing 4 oz. each, and one jumbo bar weighing 8 oz. What was the mean and median weight of his candy?

### 5. When to use the mean & standard deviation versus when to use the median & IQR

Ex:

Here are the number of games Wayne Gretsky played in each of his 20 seasons with the NHL: 79, 80, 80, 80, 74, 80, 80, 79, 64, 78, 73, 78, 74, 45, 81, 48, 80, 82, 82, 70

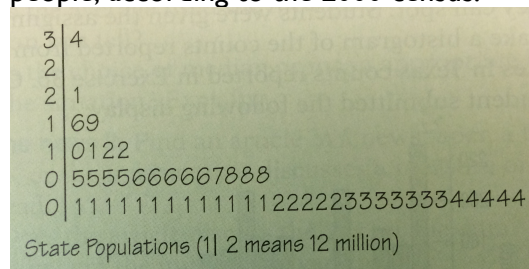
Here are the number of games Alex Ovechkin played in each of his 11 seasons thus far:

81, 82, 82, 79, 72, 79, 78, 48, 78, 81, 79

Without doing any calculations compare the mean and standard deviation of the distribution of games for each player.

- Which measure of center and spread should be used to describe each player's number of games played?
- Would it be fair to compare Gretsky's mean number of games played with Ovechkin's mean number of games played? Why or why not? Would it be fair to compare their medians? Why or why not?

The stem-and-leaf plot shows populations of the 50 states and Washington, DC, in millions of people, according to the 2000 census.



## Statistics

### Final Exam Review Topics

- What measures of center and spread are most appropriate?
- Without doing any calculations, which must be larger: the mean or the median? Explain how you know.

#### 6. Reading 2-Way tables & determining independence

Ex:

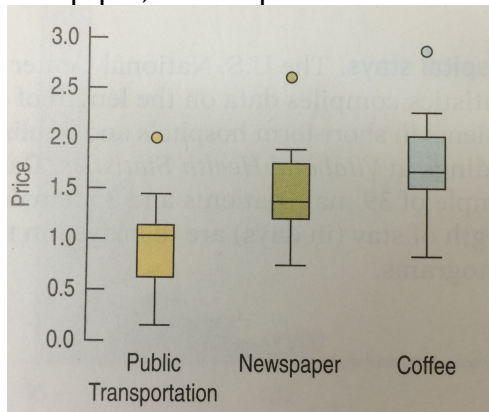
- What is the marginal distribution for gender in the following table?
- What is the conditional distribution of response for each gender?
- Are “sex” and “response” independent?
- What percent of females watch commercials?
- What percent of commercial watchers are female?
- What percent of people polled are females who watch commercials?

Response	Sex		
	Male	Female	Total
Game	279	200	479
Commercials	81	156	237
Won't watch	132	160	292
Total	492	516	1008

#### 7. Making & interpreting boxplots

Ex:

To help travelers know what to expect, researchers collected the prices of commodities in 16 cities throughout the world. Here are boxplots comparing the prices of a ride on public transportation, a newspaper, and a cup of coffee in the 16 cities (prices are all in \$US).



- Describe the distributions using SOCS.
- On average, which commodity is the most expensive? Does the presence of an outlier affect your conclusion?
- Is a newspaper always more expensive than a ride on public transportation?

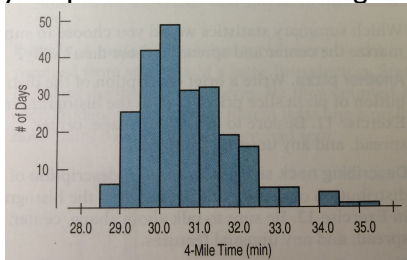
#### 8. Making & interpreting histograms

Ex:

## Statistics

### Final Exam Review Topics

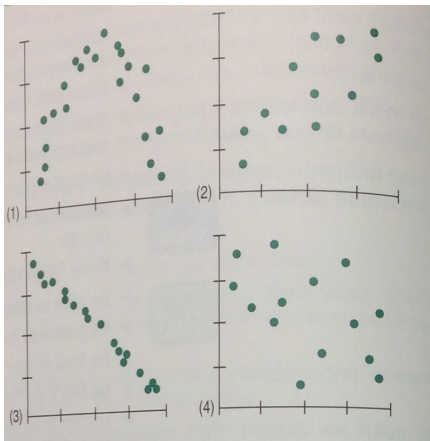
- A runner collected the times (in minutes it took him to run 4 miles on various courses during a 10-year period. Here is a histogram of the times:



- Describe the histogram using SOCS
- Would you expect the mean to be greater than or less than the median? Why?
- Estimate the mean by examining the histogram.
- Estimate the median by examining the histogram.
- Would you be surprised if the runner took 33 minutes to run four miles? Why or why not?
- Which would you use to describe this histogram, and why: the mean and standard deviation or median and IQR?
- Suppose the runner ran a 4-mile race with a PR of 27:45. Which measure(s) would this change: mean, standard deviation, median, IQR?
- During his 20 seasons in the NHL, Wayne Gretzky scored 50% more points than anyone who ever played professional hockey. Here are the number of games Gretzky played during each season. Use these data to create a histogram.  
79, 80, 80, 80, 74, 80, 80, 79, 64, 78, 73, 78, 74, 45, 81, 48, 80, 82, 82, 70

### 9. Making scatterplots, drawing & interpreting the LSRL, interpreting correlation

Ex:



- Which of the scatterplots show
  - little or no association?
  - a negative association?
  - a linear association?
  - A moderately strong association?
  - A very strong association?
  - A correlation of -0.923?
  - A correlation of -0.487?
  - A correlation of 0.006?
  - A correlation of 0.777?

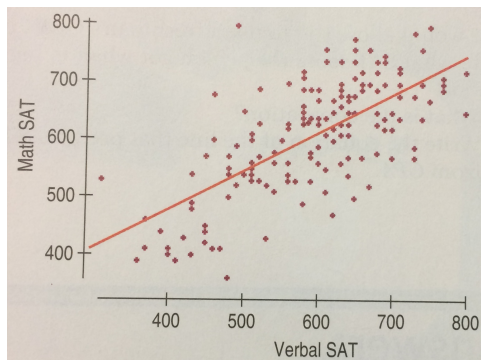


## Statistics

### Final Exam Review Topics

- Suppose you were to collect data for each pair of variables. You want to make a scatterplot. Which variable would you use as the explanatory variable and which as the response variable? What would you expect to see in the scatterplot? Describe the likely direction, form, and strength.
  - Apples: weight in grams, weight in ounces
  - Apples, circumference in inches, weight in ounces
  - College freshmen: shoe size, GPA
  - Gasoline: number of miles you drove since filling up, gallons remaining in your tank

The SAT is a test often used as part of an application to college. SAT scores are between 200 and 800 but have no units. Tests are given in both Math and Verbal areas. Doing the SAT Math problems also involves the ability to read and understand the questions, but can a person's verbal score be used to predict the math score? Verbal and math SAT scores of a high school graduating class are displayed in the scatterplot, with the regression line added. For these data  $r = 0.685$ . The equation of the least squares regression line is  $\hat{y} = 217.69 + 0.661x$



- Is a linear model appropriate?
- Interpret the slope of this regression line.
- Interpret the y-intercept of this regression line.
- Predict the math score of a student with a verbal score of 500.
- Every year, someone scores a perfect 1600. Based on this model, what would be that student's Math score residual?

### 10. Finding a z-score & finding a percentile using a z-score

Ex:

- What is a z-score? What does it tell you?
- What is the relationship between a z-score and a percentile (or percentile rank)?
- Suppose the height of horses is Normally distributed with a mean of 60.8 inches and a standard deviation of 3.2 inches. A horse on a farm measures 62 inches tall. What is that horse's z-score?
- What percent of horses are *taller* than this horse that is 62 inches tall?
- What percent of horses are between 60 and 63 inches tall?
- What can you tell about a horse that has a z-score of -0.52?

### 11. Finding a value when a percentile is known

Ex:

- Suppose the height of horses is Normally distributed with a mean of 60.8 inches and a standard deviation of 3.2 inches. What is the height of a horse in the 70<sup>th</sup> percentile of height?
- What is the height of a horse in the 40<sup>th</sup> percentile for height?

## **I 2. What is the Normal distribution**

Ex:

- What are some examples of data that would be Normally distributed?
- What does the shape of a Normal distribution look like?
- What percent of data is 1 standard deviation away from the mean in a Normal distribution? 2 standard deviations away? 3 standard deviations away?

## **I 3. Sampling & Experimental design**

Ex:

- Describe and tell situations in which you would use each type of sample: simple random sample, stratified sample, and cluster sample
- Why do we use a placebo in an experiment?
- What is the difference between an observational study and an experiment?
- Design an experiment that would help you determine which type of desk helps students write more efficiently: a flat desk like the ones we use or a slanted desk (like the lectern at the front of the room)
- What is the only way to determine causation?
- Why can we not say that putting “Slow, Children Playing” signs *causes* people to slow down on a residential street (assuming that once the signs are up, people drive more slowly)?